A New Class of Broyden Families for Nonlinear Least Squares Problems(Fundamental Technologies in Numerical Computation)

Yabe, Hiroshi

数理解析研究所講究録
第 832 巻 1993 年 137-150

A New Class of Broyden Families for Nonlinear Least Squares Problems
(非線形最小 2 乗問題に対する新しいクラスの B r o y d e n 族について)

東京理科大学・工学部　　矢部　博　(Hiroshi Yabe)

1 Introduction

This paper is concerned with the nonlinear least squares problem

$$\text{(1.1)} \qquad \text{minimize } f(x) = \frac{1}{2}\|r(x)\|^2 = \frac{1}{2}\sum_{j=1}^{m}(r_j(x))^2,$$

where $r_j : R^n \to R$, $j = 1, ..., m\,(m \geq n)$ are twice continuously differentiable, $r(x) = (r_1(x), \ldots, r_m(x))^T$ and $\| \ \|$ denotes the 2 norm. Among many numerical methods, structured quasi-Newton methods seem very promising. These methods use the structure of the Hessian matrix of $f(x)$,

$$\text{(1.2)} \qquad \nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x),$$

where J is the Jacobian matrix of r, and approximate the second part of the Hessian by some matrix A. The structured quasi-Newton methods were proposed in order to overcome the poor performance of the Gauss-Newton method for large residual problems[2],[5]. In this paper, we consider the line search strategy as a globalization technique. This generates the sequence $\{x_k\}$ by

$$x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a step length and a search direction d_k is given by solving the linear system of equations

$$\text{(1.3)} \qquad (J_k^T J_k + A_k)d = -J_k^T r_k,$$

where $r_k = r(x_k)$, $J_k = J(x_k)$, and the $n \times n$ matrix A_k is the approximation to the second part of the Hessian matrix. The matrix A_k is generated by some quasi-Newton updating formula, say, A-update. This system corresponds to the Newton equation. Since the coefficient matrix of (1.3) does not necessarily possess the hereditary positive definiteness property, Yabe and Takahashi[14] proposed computing the search direction d_k by solving the linear system of equations

$$(J_k + L_k)^T(J_k + L_k)d = -J_k^T r_k, \tag{1.4}$$

where the matrix L_k is an $m \times n$ correction matrix to the Jacobian matrix such that $(J_k+L_k)^T(J_k+L_k)$ approximates the Hessian and is generated by some updating formula, say, L-update. Since the coefficient matrix is expressed by its factorized form, the search direction may be expected to be a descent direction for f. Following Dennis[4], we dealt with the secant condition

$$(J_{k+1} + L_{k+1})^T(J_{k+1} + L_{k+1})s_k = z_k, \tag{1.5}$$

where

$$s_k = x_{k+1} - x_k, \quad z_k = (J_{k+1} - J_k)^T r_{k+1} + J_{k+1}^T J_{k+1} s_k. \tag{1.6}$$

We call this method the factorized quasi-Newton method. Yabe and Takahashi[14] proposed BFGS-like and DFP-like updates, and Yabe and Yamaki[16] obtained a structured Broyden family for L_k that contained these updates.

On the other hand, Sheng and Zou[11] studied factorized versions of the structured quasi-Newton methods independently of us. They proposed obtaining a search direction d_k by solving the linear least squares problem

$$\text{minimize } \frac{1}{2}\| r_k + (J_k + L_k)d \|^2 \text{ with respect to } d. \tag{1.7}$$

In the case of $L_k = 0$, the above implies the Gauss-Newton model. The normal equation of (1.7) is represented by

$$(J_k + L_k)^T(J_k + L_k)d = -(J_k + L_k)^T r_k. \tag{1.8}$$

Since the vector $-L_k^T r_k$ exists in the righthand side, the above does not correspond to the Newton equation, so Sheng and Zou imposed the orthogonality condition

$$L_{k+1}^T r_{k+1} = 0 \tag{1.9}$$

on the matrix L_{k+1} in addition to the secant condition (1.5). They obtained a BFGS-like update and showed the local and q-superlinear convergence of their method. The idea of Sheng and Zou seems very interesting to us, because their update includes a feature different from our factorized updates. Further, some numerical experiments given in Yabe and Takahashi[15] suggest the efficiency of their method. Recently, Yabe[13] have obtained a general form which satisfies the secant condition (1.5) and the orthogonality condition (1.9) and proposed the SZ-Broyden family (L-update);

(SZ-Broyden family (L-update))

$$\begin{aligned} L_+ = PL \; &+ \; (1-\sqrt{\phi})\left(\frac{PNs}{s^T P^\sharp s}\right)(\tau z^\sharp - P^\sharp s)^T \\ &+ \sqrt{\phi}\, PN(\tau (P^\sharp)^{-1} z^\sharp - s)\left(\frac{z^\sharp}{s^T z^\sharp}\right)^T, \end{aligned} \tag{1.10}$$

where

$$0 \le \phi \le 1, \qquad \left[(1-\phi)\frac{s^T z^\sharp}{s^T P^\sharp s} + \phi \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{s^T z^\sharp}\right] \tau^2 = 1, \tag{1.11}$$

$$Q = \frac{r_+ r_+^T}{\|r_+\|^2}, \quad P = I - Q = I - \frac{r_+ r_+^T}{\|r_+\|^2},$$

$$N = J_+ + PL, \quad B^\sharp = N^T N = (J_+ + PL)^T (J_+ + PL),$$

$$P^\sharp = N^T P N, \quad Q^\sharp = N^T Q N = \frac{J_+^T r_+ r_+^T J_+}{\|r_+\|^2} \quad \text{and} \quad z^\sharp = z - Q^\sharp s.$$

Setting $B_+ = (J_+ + L_+)^T (J_+ + L_+)$, we have

$$\begin{aligned} B_+ &= \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T P^\sharp s) v^\sharp (v^\sharp)^T\right) + Q^\sharp \\ &= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T P^\sharp s) v^\sharp (v^\sharp)^T, \end{aligned} \tag{1.12}$$

$$v^\sharp = \frac{P^\sharp s}{s^T P^\sharp s} - \frac{z^\sharp}{s^T z^\sharp}. \tag{1.13}$$

Now we have two kinds of updates, an A-update and an L-update, each with merits and demerits. An A-update just needs an $n \times n$ symmetric square matrix and is calculated in $O(n^2)$ arithmetic cost, but the coefficient matrix in (1.3) is not necessarily positive definite for the line search strategy. On the other hand, an L-update may be expected to maintain the positive definiteness of the coefficient matrix in (1.4), but it needs an $m \times n$ rectangular matrix and is calculated in $O(mn)$ arithmetic cost. However both approaches should not compete each other, but should complement each other. By using a relationship between an A-update and an L-update, special features of an L-update can be reflected in an A-update. This is a motivation of this paper.

In Section 2, we investigate a relationship between an A-update and an L-update. By using this relation, we show that the structured Broyden family given by Yabe and Yamaki[16] corresponds to the structured secant update from the convex class proposed by Engels and Martinez[7]. We also obtain a new A-update that corresponds to the SZ-Broyden family (L-update). Section 3 deals with sizing techniques, which were first proposed by Bartholomew-Biggs[2] and Dennis et al.[5]. Finally we show some numerical experiments of Broyden-like families for A-updates in Section 4, and examine the effectiveness of sizing techniques. Throughout the paper, for simplicity, we drop the subscript k and replace the subscript $k+1$ by "+". Further $\|\ \|$ denotes the 2 norm.

2 Relation between A-Updates and L-Updates, and a new Broyden-like Family

The main subject of this section is the investigation of the relationship between A-updates and L-updates. Using this relationship, we show that the structured Broyden family given by Yabe and Yamaki[16] can be regarded as the factorized version of the structuted secant update from the convex class proposed by Engels and Martinez[7]. On the other hand, Yabe[13] proposed the SZ-Broyden family (L-update) based on the idea of Sheng and Zou. This family has a feature different from our factorized updates. An application of the relationship between A-updates and L-updates to the SZ-Broyden family (L-update) enables us to obtain a new A-update which has a feature different from the family of Engels and Martinez.

Consider the case where we do not impose the orthogonality condition $L_+^T r_+ = 0$ on the matrix L_+ for the SZ-Broyden family. In this case, we may regard $P = I$. We then have

$$N = J_+ + L, \quad Q = 0, \quad Q^\sharp = 0, \quad z^\sharp = z \quad \text{and} \quad P^\sharp = B^\sharp.$$

Then the family (1.10) reduces to the structured Broyden family given by Yabe and Yamaki[16]:

$$\begin{aligned} L_+ &= L + (1-\sqrt{\phi})\left(\frac{(J_+ + L)s}{s^T B^\sharp s}\right)(\tau z - B^\sharp s)^T \\ &\quad + \sqrt{\phi}\,(J_+ + L)(\tau (B^\sharp)^{-1} z - s)\left(\frac{z}{s^T z}\right)^T, \end{aligned} \tag{2.1}$$

$$B_+ = B^\sharp - \frac{B^\sharp s s^T B^\sharp}{s^T B^\sharp s} + \frac{z z^T}{s^T z} + \phi(s^T B^\sharp s) v v^T, \tag{2.2}$$

where

$$\tau^2 = \frac{1}{(1-\phi)\dfrac{s^T z}{s^T B^\sharp s} + \phi \dfrac{z^T (B^\sharp)^{-1} z}{s^T z}}, \tag{2.3}$$

$$B^\sharp = (J_+ + L)^T (J_+ + L) \qquad \text{and} \qquad v = \frac{B^\sharp s}{s^T B^\sharp s} - \frac{z}{s^T z}. \tag{2.4}$$

In L-updates, the matrix $(J_+ + L_+)^T(J_+ + L_+)$ is a new approximation to the Hessian matrix $\nabla^2 f(x_+)$, and in A-updates, the matrix $J_+^T J_+ + A_+$ is a new approximation to the Hessian. Furthermore, the matrices $(J_+ + L)^T(J_+ + L)$ and $J_+^T J_+ + A$ are intermediate matrices for L-updates and A-updates, respectively. Thus we can regard the matrices $(J_+ + L)^T(J_+ + L)$ and $(J_+ + L_+)^T(J_+ + L_+)$ as the matrices $J_+^T J_+ + A$ and $J_+^T J_+ + A_+$, respectively. So, setting

$$B^\sharp = J_+^T J_+ + A \quad \text{and} \quad B_+ = J_+^T J_+ + A_+ \tag{2.5}$$

in (2.2), we obtain an A-update:

$$A_+ = A - \frac{w w^T}{s^T w} + \frac{z z^T}{s^T z} + \phi(s^T w) v v^T, \tag{2.6}$$

$$v = \frac{w}{s^T w} - \frac{z}{s^T z}, \qquad w = (J_+^T J_+ + A)s, \qquad 0 \le \phi \le 1,$$

which corresponds to the structured secant update from the convex class proposed by Engels and Martinez[7]. Thus the expression (2.1) can be regarded as the factorized form of their family. Note that for $\phi = 0$ and $\phi = 1$ the above implies the BFGS update of Al-Baali and Fletcher[1],[6] and the DFP update of Dennis-Gay-Welsch[5], respectively.

We have stated the relationship between A-updates and L-updates above. Now we are interested in what A-update corresponds to the SZ-Broyden family (L-update), so we apply the relation (2.5) to the SZ-Broyden family (1.12). Since $Q^\sharp = J_+^T Q J_+$ and $Q = r_+ r_+^T / \|r_+\|^2$, we have

$$P^\sharp s = (B^\sharp - Q^\sharp)s = As + J_+^T(I - Q)J_+ s$$

and

$$z^\sharp = z - Q^\sharp s = (J_+ - J)^T r_+ + J_+^T(I - Q)J_+ s.$$

Thus we obtain a new A-update:
(SZ-Broyden family (A-update))

$$A_+ = A - \frac{w^\sharp (w^\sharp)^T}{s^T w^\sharp} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T w^\sharp) v^\sharp (v^\sharp)^T, \tag{2.7}$$

where

$$\begin{aligned} v^\sharp &= \frac{w^\sharp}{s^T w^\sharp} - \frac{z^\sharp}{s^T z^\sharp}, \qquad w^\sharp = As + J_+^T \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) J_+ s, \\ z^\sharp &= (J_+ - J)^T r_+ + J_+^T \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) J_+ s \qquad \text{and} \qquad 0 \le \phi \le 1. \end{aligned}$$

For $\phi = 0$ and $\phi = 1$, we have an SZ-BFGS update (A-update) and an SZ-DFP update (A-update), respectively:

$$A_+ = A - \frac{w^\sharp (w^\sharp)^T}{s^T w^\sharp} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}$$

and

$$A_+ = A - \frac{w^\sharp (z^\sharp)^T + z^\sharp (w^\sharp)^T}{s^T z^\sharp} + \left(1 + \frac{s^T w^\sharp}{s^T z^\sharp} \right) \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}.$$

Note that the preceding updates contain the projection information of the orthogonality condition (1.9) for an L-update in the vectors $w^\sharp$ and $z^\sharp$. Thus we may expect this new A-update to possess a feature different from the family of Engels and Martinez in practical computations.

3 Sizing Techniques

We know that, for large residual problems, the sructured quasi-Newton methods perform well, but that for zero and small residual problems, the Gauss-Newton method performs better. Thus, in the latter case, it is desirable for the structured quasi-Newton methods to follow the Gauss-Newton method. For this purpose, Bartholomew-Biggs[2] and Dennis et al.[5] introduced sizing techniques, and Al-Baali et al.[1] considered the combination of the structured quasi-Newton methods and the Gauss-Newton method — hybrid methods.

Though both strategies can be applied to all the methods given in the previous sections, we consider only sizing techniques in what follows.

For A-updates, Bartholomew-Biggs[2] proposed a sizing parameter (Biggs parameter)

$$\beta_k = \frac{r(x_{k+1})^T r(x_k)}{r(x_k)^T r(x_k)} \tag{3.1}$$

based on the idea such that if $r(x_{k+1}) = \beta_k r(x_k)$ for some β_k, $A_k = \sum_{i=1}^m r_i(x_k)\nabla^2 r_i(x_k)$ and each $r_i(x_k)$ is quadratic, then $\sum_{i=1}^m r_i(x_{k+1})\nabla^2 r_i(x_{k+1}) = \beta_k A_k$. Dennis et al.[5] proposed a sizing parameter (DGW parameter)

$$\beta_k = \min\left(\left|\frac{s_k^T(J_{k+1} - J_k)^T r_{k+1}}{s_k^T A_k s_k}\right|, 1\right) \tag{3.2}$$

based on the idea that the spectrum of the sized matrix $\beta_k A_k$ should overlap that of the second part of the Hessian matrix in the direction of s_k. Note that the factor $s_k^T(J_{k+1} - J_k)^T r_{k+1}/s_k^T A_k s_k$ corresponds to the factor given by Oren[9].

Now we present an algorithm for structured quasi-Newton methods with sizing techniques.

(Algorithm A for A-updates)
Starting with a point $x_1 \in R^n$ and an $n \times n$ matrix A_1 (usually, $A_1 = 0$ and $\beta_1 = 1$), the algorithm proceeds, for $k = 1, 2, \ldots$, as follows:

Step 1. Having x_k and A_k, find the search direction d_k by solving the linear system of equations

$$(J_k^T J_k + A_k)\, d = -J_k^T r_k. \tag{3.3}$$

Step 2. Choose a steplength α_k by a suitable line search algorithm.

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. If the new point satisfies the convergence criterion, then stop; otherwise, go to Step 5.

Step 5. Construct A_{k+1} by using the following A-updates:

(Engels and Martinez family)

$$A_{k+1} = \beta_k A_k - \frac{w_k w_k^T}{s_k^T w_k} + \frac{z_k z_k^T}{s_k^T z_k} + \phi_k (s_k^T w_k) v_k v_k^T, \tag{3.4}$$

where

$$v_k = \frac{w_k}{s_k^T w_k} - \frac{z_k}{s_k^T z_k}, \quad w_k = (J_{k+1}^T J_{k+1} + \beta_k A_k) s_k,$$

or

(SZ-Broyden family (A-update))

$$A_{k+1}=\beta_k A_k-\frac{w_k^\sharp(w_k^\sharp)^T}{s_k^T w_k^\sharp}+\frac{z_k^\sharp(z_k^\sharp)^T}{s_k^T z_k^\sharp}+\phi_k(s_k^T w_k^\sharp)v_k^\sharp(v_k^\sharp)^T, \tag{3.5}$$

where

$$\begin{aligned} v_k^\sharp &= \frac{w_k^\sharp}{s_k^T w_k^\sharp}-\frac{z_k^\sharp}{s_k^T z_k^\sharp}, \\ w_k^\sharp &= \beta_k A_k s_k + J_{k+1}^T(I-(\|r_{k+1}\|^2)^\dagger r_{k+1}r_{k+1}^T)J_{k+1}s_k, \\ z_k^\sharp &= (J_{k+1}-J_k)^T r_{k+1} + J_{k+1}^T(I-(\|r_{k+1}\|^2)^\dagger r_{k+1}r_{k+1}^T)J_{k+1}s_k, \end{aligned}$$

and β_k is defined by the Biggs parameter (3.1) or the DGW parameter (3.2), ϕ_k is a parameter such that

$$0 \le \phi_k \le 1,$$

and $(\|r_{k+1}\|^2)^\dagger$ denotes the Moore-Penrose generalized inverse of $\|r_{k+1}\|^2$.

4 Computational Experiments

The purposes of our numerical experiments are to compare a new A-update (3.5) with the Engels and Martinez family (3.4), and to investigate how the computational performance depends on the choice of the parameters ϕ_k and β_k given in Algorithm A from the viewpoint of the number of iterations and the number of vector valued function (i.e. $r(x)$) evaluations. Note that there are different strategies among the structured quasi-Newton methods. Dennis et al.[5] combined the DGW update and the trust region globalization strategy, and Al-Baali et al.[1] proposed the hybrid method which combined the Gauss-Newton method and the structured BFGS update for the line search globalization strategy, and so forth. In this section, we just compare the performance of some updates for the line search globalization strategy.

The numerical calculations were carried out in double precision arithmetic on a SUN SPARC station 1+, and the program was coded in FORTRAN 77. The Jacobian matrix is evaluated by the forward difference approximation. In Algorithm A, the initial matrix A_1 is set to zero matrix. The linear system of equations in Step 1 is solved by the modified Cholesky method, i.e. when the coefficient matrix cannot be decomposed because of indefiniteness, a diagonal element of Cholesky factor was replaced by a small positive number. In Step 2, the bisection line search method with Armijo's rule

$$f(x_k+\alpha_k d_k) \le f(x_k) + 0.1\alpha_k \nabla f(x_k)^T d_k \tag{4.1}$$

is employed. Further, in Step 4, the iterative process is terminated

(T1) if $\|r(x_{k+1})\|_\infty \le \max(\mathrm{TOL1},\, \varepsilon)$, or

(T2) if $|e_i^T J(x_{k+1})^T r(x_{k+1})| \le \max(\mathrm{TOL2}, \varepsilon)\|r(x_{k+1})\|\|J(x_{k+1})e_i\|$ for $i = 1, \ldots, n$ and $\|x_{k+1} - x_k\|_\infty \le \max(\mathrm{TOL3}, \varepsilon) \max(\|x_{k+1}\|_\infty, 1.0)$, where e_i denotes the i-th column of the unit matrix,

or

(T3) if the number of iterations exceeds the prescribed limit (ITMAX),

or

(T4) if the number of function evaluations exceeds the prescribed limit (NFEMAX),

where $\|\bullet\|_\infty$ denotes the maximum norm and ε is machine epsilon. The modified Cholesky method and the stopping criteria described above followed the code NOLLS1 in Tanabe and Ueda[12]. In the experiments, we set TOL1 = TOL2 = TOL3 = 10^{-4}, ITMAX = 500 and NFEMAX = 2000. For the SZ-Broyden family (A-update) in Step 5, the Moor-Penrose generalized inverse $(\|r_{k+1}\|^2)^\dagger$ was numerically set to $(1/\|r_{k+1}\|^2)$ if $\|r_{k+1}\|^2 \ge 10^{-20}$, and 0 otherwise. Since the stopping criteria (T1) with TOL1 = 10^{-4} was used, $(\|r_{k+1}\|^2)^\dagger$ was not set to zero in our numerical experiments. In addition to the convex classes of the Broyden-like families mentioned in the previous sections, we used the Gauss-Newton method (GN) and the structured symmetric rank one (SR1) update for comparison. The structured SR1 update with sizing was first proposed by Bartholomew-Biggs[2], and is represented by

$$A_{k+1} = \beta_k A_k + \frac{(q_k - \beta_k A_k s_k)(q_k - \beta_k A_k s_k)^T}{s_k^T (q_k - \beta_k A_k s_k)^T}, \tag{4.2}$$

where

$$q_k = (J_{k+1} - J_k)^T r_{k+1}.$$

Since the DGW sizing parameter (3.2) makes the denominator zero, we just applied the Biggs sizing parameter to the above.

The names, the sizes and the starting points of the test problems we used are listed in Table 1, together with the abbreviated problem names. These problems are given in Dennis et al.[5], and are in detail in Moré, Garbow and Hillstrom[8]. In Table 1, (Z), (S) and (L) mean a zero residual problem, a small residual problem and a large residual problem, respectively. Tables 2 and 5 are computational results for the Engels and Martinez family with no sizing, with DGW sizing and Biggs sizing parameters, respectively. Tables 3 and 6 are for the SZ-Broyden (A-update) with no sizing, DGW sizing and Biggs sizing parameters, respectively. The computational results for the Gauss-Newton method and the structured SR1 update are given in Tables 4 and 7. In each table, the total number of iterations and the total number of function evaluations are written. The latter is written in a parenthesis in the tables, and contains the number to evaluate the Jacobian matrix by forward finite difference. Also, the asterisk $*$ in each table contains the case where the method failed to converge in the specified number of iterations or function evaluations. In each table, the number in the parenthesis denotes the performance ratio of sizing techniques. For example, in the part of B(0.1) with DGW sizing in Table 2, the ratio 0.844 implies 304/360. The small ratio means that the sizing technique works very well. However we should note that this ratio depends on the choices of ITMAX and NFEMAX in the stopping criteria in the case where the symbol $*$ is attached. In each table, "B(ξ)" means the results of the Broyden-like family with $\phi_k = \xi$. So, in the Engels and Martinez family, "B(0.0)" and "B(1.0)" corresponds to the results of the Al-Baali and Fletcher update and the revised DGW update, respectively. However, since Al-Baali and Fletcher proposed the hybrid method and Dennis et al. used the trust region strategy, we cannot make a direct comparison with their results.

From all the tables, we summarize our numerical results as follows:
(1)The structured quasi-Newton methods with sizing are more robust than the Gauss-Newton method.
(2)The Engels and Martinez family matched with the DGW sizing parameter, and the SZ-Broyden family (A-update) matched with the Biggs sizing parameter. Further the SZ-Broyden family with Biggs sizing parameter worked better than the other families.
(3)For both families with sizing, the cases of $\phi = 0.5$ were numerically stable.
(4)As the parameter ϕ approached 1, the performance of sizing techniques increased. The DFP-like update without sizing was much inferior to other updates without sizing. On the other hand, the DFP-like update with sizing worked as well as the other sized updates.
(5)The Bartholomew-Biggs update, i.e. the structured symmetric rank one update with Biggs parameter, worked well.

The result (2) suggests that an application of features of L-updates to A-updates is promissing. In this paper, we suggested one relationship between A-updates and L-updates. This result encouraged us to study another relation and to propose a new A-update which corresponds to a new L-update. The result (3) is somewhat similar to the numerical results given in Oren[10], in which Oren applied his sizing parameter to the standard Broyden family for general minimization problems. The result (4) means that the DFP-like update needs sizing technique very much. However this does not mean that the other updates, e.g. BFGS-like update, need no sizing. There is a possibility of finding another kind of sizing parameter which is effective for the other updates. This result supports the research of Contreras and Tapia[3]. In their paper, they claimed that the standard DFP update needed to be sized for general minimization problems, and that the DFP update was much imposed when matched with the Oren-Luenberger sizing parameter. They proposed another kind of sizing parameter for the standard BFGS update. Their idea may be applied to the structured quasi-Newton methods. The result (5) encourages us to study a nonconvex class of the structured Broyden family, because the structured symmetric rank one update does not belong to the convex class but is a member of the Broyden-like family.

5 Concluding Remarks

The numerical results show that the SZ-Broyden family (A-update) with the Biggs sizing parameter works well. These results also show that the DFP-like update needs sizing techniques very much and supports the research of Contreras and Tapia[3]. Further investigation of the relationship between A-updates and L-updates seems very promising to us. This may give us a new A-updates. Since L-updates enable us to obtain a descent search direction for the objective function, by investigating the relation we may expect to find conditions under which matrices $J_k^T J_k + A_k$ possess the hereditary positive definiteness property for A-updates. However, the relation mensioned in this paper is not exact yet, because the intermediate matrices of A-updates and L-updates do not in general correspond exactly. The results of Section 2 seem to give us a clue to understanding the relation.

This paper mainly dealt with the convex classes of the Broyden-like families. As mentioned in the previous section, updates which are not contained in the convex classes are also promising. The structured SR1 update is especially interesting. In fact, setting

$$\phi_k = \frac{s_k^T z_k}{s_k^T (z_k - w_k)} \qquad \text{and} \qquad \phi_k = \frac{s_k^T z_k^\sharp}{s_k^T (z_k^\sharp - w_k^\sharp)}$$

in (3.4) and (3.5), respectively, we have

$$A_{k+1} = \beta_k A_k + \frac{(z_k - w_k)(z_k - w_k)^T}{s_k^T (z_k - w_k)}$$

and

$$A_{k+1} = \beta_k A_k + \frac{(z_k^\sharp - w_k^\sharp)(z_k^\sharp - w_k^\sharp)^T}{s_k^T (z_k^\sharp - w_k^\sharp)}.$$

Since $z_k - w_k = z_k^\sharp - w_k^\sharp = (J_{k+1} - J_k)^T r_{k+1} - \beta_k A_k s_k$, the above yields the structured SR1 update (4.2). This means that the structured SR1 update is a common member of the nonconvex classes of the Engels-Martinez family and the SZ-Broyden family (A-update). However, note that the projection information of the orthogonality condition (1.9) is no longer included in the structured SR1 update.

References

1. M.Al-Baali and R.Fletcher (1985) Variational methods for non-linear least squares. *Journal of the Operational Research Society* 36 , No.5, 405-421.

2. M.C.Bartholomew-Biggs (1977) The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals. *Mathematical Programming* 12, 67-80.

3. M.Contreras and R.A.Tapia (1991) *Sizing the BFGS and DFP updates: A numerical study*, Technical Report TR91-19, July, Dept. of Mathematical Sciences, Rice University, Houston, Texas, USA.

4. J.E.Dennis,Jr. (1976) A brief survey of convergence results for quasi-Newton methods. *SIAM-AMS Proceedings* 9, 185-199.

5. J.E.Dennis,Jr., D.M.Gay and R.E.Welsch (1981) An adaptive nonlinear least squares algorithm. *ACM Transactions on Mathematical Software* 7, No.3, 348-368.

6. J.E.Dennis,Jr., H.J.Martinez and R.A.Tapia (1989) Convergence theory for the structured BFGS secant method with an application to nonlinear least squares. *Journal of Optimization Theory and Applications* 61, No.2, 161-178.

7. J.R.Engels and H.J.Martinez (1991) Local and superlinear convergence for partially known quasi-Newton methods. *SIAM J. on Optimization* 1, No.1, 42-56.

8. J.J.Moré, B.S.Garbow and K.E.Hillstrom (1981) Testing unconstrained optimization software. *ACM Transactions on Mathematical Software* 7, No.1, 17-41.

9. S.S.Oren (1974) Self-scaling variable metric (SSVM) algorithms, Part 2: Implementation and Experiments. *Management Science* 20, No.5, 863-874.

10. S.S.Oren (1974) On the selection of parameters in self scaling variable metric algorithms. *Mathematical Programming* 7, 351-367.

11. S.Sheng and Z.Zou (1988) *A new secant method for nonlinear least squares problems*, Technical Report NANOG-1988-03, March, Nanjing University, People's Republic of China.

12. K.Tanabe and S.Ueda (1981) NOLLS1, A FORTRAN subroutine for nonlinear least squares by a quasi-Newton method. *Computer Science Monographs* 17, The Institute of Statistical Mathematics, Tokyo, Japan.

13. H.Yabe (1991) Variations of structured Broyden families for nonlinear least squares problems. in *Numerical Analysis and Scientific Computation*, Research Report 746, pp.156-165, Research Institute for Mathematical Sciences, Kyoto University, Japan.

14. H.Yabe and T.Takahashi (1991) Factorized quasi-Newton methods for nonlinear least squares problems. *Mathematical Programming* 51, No.1, 75-100.

15. H.Yabe and T.Takahashi (1991) Numerical comparison among structured quasi-Newton methods for nonlinear least squares problems. *Journal of the Operations Research Society of Japan* 34, No.3, 287-305.

16. H.Yabe and N.Yamaki (1991) Convergence of structured quasi-Newton methods with structured Broyden family, In *Nonlinear Optimization - Modeling and Algorithms*. Cooperative Research Report 29, March, The Institute of Statistical Mathematics, Tokyo, Japan, *in Japanese*, 120-136.

Table 1. Test Problems

Abbrebiated Name	Name of Test Problem	m	n	Starting Point	Residual
WATSON6	Watson Problem with 6 variables	31	6	(0, 0, ..., 0)	(S)
WATSON9	Watson Problem with 9 variables	31	9	(0, 0, ..., 0)	(S)
WATSON12	Watson Problem with 12 variables	31	12	(0, 0, ..., 0)	(S)
WATSON20	Watson Problem with 20 variables	31	20	(0, 0, ..., 0)	(S)
ROSENBROCK	Rosenbrock Problem	2	2	(-1.2, 1.0)	(Z)
HELIX	Helical Valley Problem	3	3	(-1, 0, 0)	(Z)
POWELL	Powell's Singular Problem	4	4	(3, -1, 0, 1)	(Z)
BEALE	Beale Problem	3	2	(0.1, 0.1)	(Z)
FRDSTEIN1	Freudenstein and Roth Problem	2	2	(6, 6)	(Z)
FRDSTEIN2	Freudenstein and Roth Problem	2	2	(15, -2)	(L)
BARD	Bard Problem	15	3	(1, 1, 1)	(S)
BOX	Box Problem	10	3	(0, 10, 20)	(Z)
KOWALIK	Kowalik Problem	11	4	(0.25 , 0.39, 0.415, 0.39)	(S)
OSBORNE1	Osborne Problem	33	5	(0.5, 1.5 , -1.0, 0.01, 0.02)	(S)
OSBORNE2	Osborne Problem	65	11	(1.3, 0.65, 0.65, 0.7, 0.6 , 3.0 , 5.0, 7.0 , 2.0 , 4.5 , 5.5)	(S)
JENNRICH	Jennrich Problem	10	2	(0.3, 0.4)	(L)

Table 2 Total Number of Iterations (Engels and Martinez family)

	BFGS B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No Sizing	431*	360	404	388	380
DGW Sizing	387*	304	298	300	315
	(0.898)	(0.844)	(0.738)	(0.773)	(0.829)
Biggs Sizing	325	323	333	320	318
	(0.754)	(0.897)	(0.824)	(0.825)	(0.837)

Table 2 (Continued)

	B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No Sizing	395	437	407	502*	658	1467*
DGW Sizing	300	295	288	292	286	287
	(0.759)	(0.675)	(0.708)	(0.582)	(0.435)	(0.196)
Biggs Sizing	300	302	349	362	363	312
	(0.759)	(0.691)	(0.857)	(0.721)	(0.552)	(0.213)

Table 3 Total Number of Iterations (SZ-Broyden family (A update))

	BFGS B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No Sizing	518	443	535	514	501
DGW Sizing	299	300	308	316	310
	(0.577)	(0.677)	(0.576)	(0.615)	(0.619)
Biggs Sizing	320	301	300	307	299
	(0.618)	(0.679)	(0.561)	(0.597)	(0.597)

Table 3 (Continued)

	B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No Sizing	686*	577	771	979	1190*	2819*
DGW Sizing	309	305	299	253	306	305
	(0.450)	(0.529)	(0.388)	(0.258)	(0.257)	(0.108)
Biggs Sizing	301	319	302	328	358	324
	(0.439)	(0.553)	(0.392)	(0.335)	(0.301)	(0.115)

Table 4 Total Number of Iterations (Gauss-Newton, SR1)

	GN	SR1
No Sizing	478*	316
Biggs Sizing	—	302
		(0.956)

Table 5 Total of Function Evaluations (Engels and Martinez family)

	BFGS B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No Sizing	4269*	3281	4160	3530	3189
DGW Sizing	3952*	2766	2697	2728	2754
	(0.926)	(0.843)	(0.648)	(0.773)	(0.864)
Biggs Sizing	3018	2839	2825	2833	2774
	(0.707)	(0.865)	(0.679)	(0.803)	(0.870)

Table 5 (Continued)

	B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No Sizing	3507	3833	3532	4537*	4423	7990*
DGW Sizing	2769	2724	2650	2653	2654	2629
	(0.790)	(0.711)	(0.750)	(0.585)	(0.600)	(0.329)
Biggs Sizing	2777	2671	3017	3128	3056	2756
	(0.792)	(0.697)	(0.854)	(0.689)	(0.691)	(0.345)

Table 6 Total of Function Evaluations (SZ-Broyden family (A update))

	BFGS B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No Sizing	4744	3842	4139	4121	4095
DGW Sizing	2670	2731	2831	2861	2822
	(0.563)	(0.711)	(0.684)	(0.694)	(0.689)
Biggs Sizing	2980	2650	2654	2657	2618
	(0.628)	(0.690)	(0.641)	(0.645)	(0.639)

Table 6 (Continued)

	B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No Sizing	6082*	4739	6008	6961	8814*	15996*
DGW Sizing	2824	2779	2735	2146	2721	2813
	(0.464)	(0.586)	(0.455)	(0.308)	(0.309)	(0.176)
Biggs Sizing	2655	2707	2641	2749	2946	2766
	(0.437)	(0.571)	(0.440)	(0.395)	(0.334)	(0.173)

Table 7 Total of Function Evaluations (Gauss-Newton, SR1)

	GN	SR1
No Sizing	6299*	2796
Biggs Sizing	—	2691
		(0.962)